

2. Chapin, H., Jantzen, K., Kelso, J.S., Steinberg, F., and Large, E. (2010). Dynamic emotional and neural responses to music depend on performance expression and listener experience. *PLoS One* 5, 13812.
3. Zhang, Y.S., and Ghazanfar, A.A. (2016). Perinatally influenced autonomic system fluctuations drive infant vocal sequences. *Curr. Biol.* 26, 1249–1260.
4. Marler, P. (1970). A comparative approach to vocal learning: Song development in white-crowned sparrows. *J. Comp. Phys. Psych.* 71, 2.
5. Doupe, A.J., and Kuhl, P.K. (1999). Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* 22, 567–631.
6. Jarvis, E.D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., Medina, L., Paxinos, G., Perkel, D.J., Shimizu, T., *et al.* (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nat. Rev. Neurosci.* 6, 151–159.
7. Ackermann, H., Hage, S.R., and Zeigler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behav. Brain Sci.* 37, 529–546.
8. Altmann, J., and Samuels, A. (1992). Costs of maternal care: infant-carrying in baboons. *Behav. Ecol. Sociobiol.* 29, 391–398.
9. Hiraiwa, M. (1981). Maternal and alloparental care in a troop of free-ranging Japanese monkeys. *Primates* 22, 309–329.
10. Zeskind, P.S., and Lester, B.M. (1978). Acoustic features and auditory perceptions of the cries of newborns with prenatal and perinatal complications. *Child Dev.* 49, 580–589.
11. Nathani, S., Ertmer, D.J., and Stark, R.E. (2006). Assessing vocal development in infants and toddlers. *Clin. Linguist Phon.* 20, 351–369.
12. Oller, D.K. (2000). *The Emergence of the Speech Capacity* (Mahwah, NJ: Lawrence Erlbaum and Associates).
13. Papoušek, M. (1994). *Vom ersten Schrei zum ersten Wort: Anfänge der Sprachentwicklung in der vorsprachlichen Kommunikation* (Bern: Verlag Hans Huber).
14. Scheiner, E., Hammerschmidt, K., Jürgens, U., and Zwirner, P. (2002). Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of infants. *J. Voice* 16, 509–529.
15. Oller, D.K., Buder, E.H., Ramsdell, H.L., Warlaumont, A.S., Chorna, L., and Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proc. Natl. Acad. Sci. USA* 110, 6318–6323.

Animal Evolution: Are Phyla Real?

Andreas Hejnol¹ and Casey W. Dunn²

¹Sars International Centre for Marine Molecular Biology, University of Bergen, 5008 Bergen, Norway

²Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912, USA

Correspondence: andreas.hejnol@uib.no (A.H.), casey_dunn@brown.edu (C.W.D.)

<http://dx.doi.org/10.1016/j.cub.2016.03.058>

A recent study compares developmental gene expression among very distantly related animal species and interprets the results as providing biological justification for phyla. Several problems in the design and interpretation of the study call this conclusion into question.

The seeds of modern efforts to understand biodiversity were sewn by the botanist Carl Linnaeus [1]. He had the critical insight that the diversity of life is hierarchically nested in groups that are each characterized by particular suites of traits. Animals with hair (mammals), for example, are a nested subset of animals with four limbs (tetrapods). Tetrapods are a nested subset of animals with vertebrae (vertebrates). This remarkable insight is even more astonishing when one recognizes that it preceded, and in many ways enabled, the understanding that this diversity was produced through evolution by common descent [2]. Evolution explains the process that gives rise to this nested pattern — the traits that characterize more inclusive groups (e.g., the four limbs of tetrapods) evolved prior to the traits that

characterize more restricted groups (e.g., the hair of mammals).

Linnaeus sought to bring additional structure to his nested taxonomy. His work preceded the concept of phylogenies — evolutionary trees that describe the relationships between organisms — which now provide a mechanism-based framework for describing the nested structure of diversity. He therefore invented ranks — uniform levels of nesting that he applied across groups of organisms — to organize his understanding of diversity. The original ranks designated by Linnaeus were kingdom, class, order, genus and species [1], and others, including the phylum [3], have been added since. As phylogenies improve, some have called for the abandonment of ranks [4] to avoid implying that different clades given the same rank,

for instance the phyla Echinodermata and Mollusca, are somehow equivalent. Advocates for the continued use of ranks have argued that they are still useful organizational aids that help convey which clades are within others, but have emphasized that ranks do not reflect distinct biological patterns or processes and have called for others to stop trying to imbue them with biological significance [5]. Even so, it is still common in the literature, as exemplified by a new paper from Levin *et al.* [6], to treat ranks as a biological pattern to be explained rather than as a human-created organizational aid.

All living animals are now placed in about 35 animal phyla. For much of their history, the animal phyla have been the maximally inclusive groups of animals that were thought to be monophyletic (i.e., include all the descendants of their

most recent common ancestor), but whose relationships to each other were difficult to evaluate. This was explicitly stated by Haeckel [3] when he first introduced phyla: “we must therefore be very careful when determining the size and content of the individual Phylae, and prefer to accept rather a larger number of phyla, which we are each sure represent organisms of common origin, than a smaller number of stems which maybe one or the other are possibly composed of originally separate stems.” Phyla were therefore established along lines of phylogenetic uncertainty rather than on unique properties that differentiate them from other clades.

There have never been clear objective criteria for designating a clade as a phylum [7]. There is, for example, no known property that is shared by all echinoderms or all molluscs, both which have been designated as phyla, and that would distinguish these clades from more or less inclusive clades that aren't phyla, such as deuterostomes or bivalves. This has provoked considerable discussion about whether the categorization into phyla says more about the way humans perceive animal diversity or the biological processes that generated this diversity [7–11]. This can be reduced to a testable empirical question: did an evolutionary process occur along the 35 stems of the phyla that is distinct from the evolutionary changes along the millions of other branches of the animal phylogeny (Figure 1A)? Such a significant, consistent difference can be called the ‘phylum hypothesis’. If there were a consistent difference, then there would be a justification for treating phyla as somehow sharing certain biological properties. No such properties have been identified to date, and phyla are very different from each other in many respects. Phylum stems, for example, do not all span the same geological interval [8], and phyla do not contain similar numbers of species, varying from one to more than a million described species [7].

Despite a lack of evidence, the phylum hypothesis is often taken for granted. Animal evolutionary developmental biology (EvoDevo) in particular has embraced phyla as real biological entities to be explained, and much work in the field seeks to understand the origin of

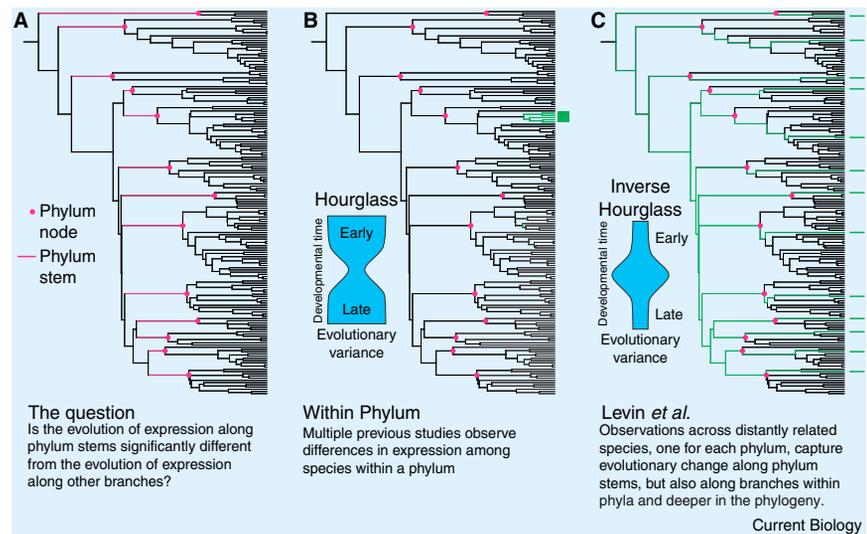


Figure 1. Simulated schematic representing the animal phylogeny.

The actual animal phylogeny has millions of tips and millions of internal branches. The most recent common ancestor of each phylum is designated with a magenta circle. (A) The phylum hypothesis postulates that the evolutionary processes along the phylum stems (magenta, shown in A only) are different than along the millions of other branches (black). (B) Previous studies of closely related species within a phylum found that the evolution of expression is more conserved in mid-development than early or late development. (C) Levin *et al.* [9] looked across 10 distantly related species, each from a different phylum, and found that the evolution of expression was less conserved in mid-development. They claimed that this is evidence for a unique process that gives rise to phyla. The primary problem with this conclusion is that their sampling strategy cannot differentiate between evolutionary processes within phyla, along phylum stems, or along other branches between phyla. The data, therefore, do not contain any information about the uniqueness of phyla, as the observed differences could have arisen along any of the green branches.

differences between phyla [10]. This is in spite of the fact that many interesting developmental differences exist both at larger scales and within phyla [12]. There have been many attempts in EvoDevo to provide a biological definition for what a phylum is, but all have been unsuccessful and circular [7]. For example, a phylum is often defined as a clade with a ‘unique body plan’ and a ‘unique body plan’ is defined as features that are specific to a phylum [10].

Based on a comparative transcriptome survey of the development of ten distantly related animal species, each from a different phylum, Levin *et al.* [6] recently argued that they have found long-sought biological criteria that objectively define phyla. Earlier studies of closely related species of fruit flies [13] and of jawed animals [14] had found that gene expression is more conserved at the mid-phase of development than early and late in development (Figure 1B). By contrast, Levin *et al.* [9] found that between the representatives of the ten phyla gene expression is less conserved in the mid-phase of development than early and late

in development (Figure 1C), the opposite of the pattern seen when comparing closely related species. From this result, they conclude that the designation of phyla is biologically justified in that it corresponds to specific evolutionary changes in developmental gene expression. In addition, Levin *et al.* [9] conclude that a phylum may be objectively defined as a collection of species whose gene expression at the mid-term phase of development is both highly-conserved among them, yet divergent relative to species of a different phylum.

There are, however, problems with these conclusions. The biggest problem is that the data Levin *et al.* [9] collected, from a single representative of each of ten phyla, are not adequate to test this definition of phyla. With millions of living species at its tips, the animal phylogeny has millions of internal branches. Only 35 of these are phylum stems (Figure 1A). Previous studies focused on closely related taxa that only captured evolutionary changes in expression along branches within phyla (Figure 1B), but

were not designed to capture differences among phyla [13,14]. Levin *et al.* [9] sampled one species from each of ten phyla. Differences between these species capture evolutionary changes in expression that have occurred within phyla, along phylum stems, and along deeper branches between phyla (Figure 1C). So while their sampling can capture changes in expression that occur along phylum stems, it is not sufficient to identify which changes are specific to phylum stems. Much denser species sampling that at least brackets the root of each phylum would be needed to identify such changes and test the phylum hypothesis.

Another problem is that Levin *et al.* [9] relied on pairwise comparisons that treat each of the ten species as an independent observation. This is statistically problematic, as some sampled species are more closely related and therefore are expected to be more similar to each other — they are not independent observations [15]. As a result of this phylogenetic structure, some changes will be over-represented in pairwise comparisons relative to other changes. For example, changes in expression along the branch that gave rise to bilaterally symmetric animals (bilaterians) will impact all the comparisons between bilaterians and non-bilaterians, even though each of these changes occurred only once. Phylogenetic comparative methods, such as ancestral character state reconstructions and independent contrasts [15], address exactly these issues [16] but were not employed. Other evolutionary analyses of gene expression, including the approaches called ‘phylostratigraphy’ [17], also do not adequately account for phylogenetic structure in the data. Not only does this mislead the analyses that have been done, it is a lost opportunity to gain an explicit understanding of the evolutionary history that led to the diversity among living species. The lack of explicit phylogenetic approaches has become a major problem in EvoDevo as a whole, not just the analyses of developmental gene expression. The evolution of development and gene expression should be studied with the same phylogenetic comparative tools that are used to study the evolution of

other traits [18]. This will avoid well-known statistical problems and enable more insightful analyses.

Nonetheless, even if there is not evidence that these differences are specific to the origins of phyla, the main finding of Levin *et al.* [9] that gene expression in the mid-phase of development could be less conserved at large evolutionary scales than gene expression early and late in development is very interesting. As only relatively close and very distantly related animal species have been compared so far, the key for understanding these differences lies in the extension of the species sampling to obtain a higher resolution understanding of which gene expression changes occurred along which branches (Figure 1).

After 150 years with phyla at the center of studies of animal diversity, there is still no support for the phylum hypothesis. This lack of support, despite many efforts to find it, makes it unproductive to continue to use phyla to frame central questions in animal evolution and development. Moreover, the value of all ranks, including phyla, as organizational tools has been supplanted by well-resolved phylogenies that provide explicit, objective tools for describing diversity and testing biological hypotheses across the entire tree of life. As long as we present animal diversity with phylum handles, biologists will be tempted to focus on the handles instead of the biology at hand. Rather than continue to seek to explain phyla in the absence of any evidence that they originated by distinct evolutionary processes, we should use general phylogenetic approaches to trace trait evolution. This will allow us to describe and explain changes anywhere in the phylogeny, not just along 35 of the millions of branches in the animal tree of life.

REFERENCES

1. Linnaeus, C. (1758). *Systema Naturæ*, 10th Edition (Stockholm: Laurentius Salvius).
2. Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (London: John Murray).
3. Haeckel, E. (1866). *Generelle Morphologie der Organismen* (Berlin: Verlag von Georg Reimer).
4. de Queiroz, K., and Gauthier, J. (1994). Toward a phylogenetic system of biological nomenclature. *Trends Ecol. Evol.* 9, 27–31.
5. Giribet, G., Hormiga, G., and Edgecombe, G. (2016). The meaning of categorical ranks in evolutionary biology. *Org. Divers. Evol.*, in press.
6. Levin, M., Anavy, L., Cole, A., Winter, E., Mostov, N., Khair, S., Senderovich, N., Kovalev, E., Silver, D., Feder, M., *et al.* (2016). The mid-developmental transition and the evolution of animal body plans. *Nature* 537, 637–641.
7. Scholtz, G. (2004). *Baupläne versus ground patterns, phyla versus monophyla: aspects of patterns and processes in evolutionary developmental biology*. In *Evolutionary Developmental Biology of Crustacea*, G. Scholtz, ed. (Lisse: A.A. Balkema Publishers), pp. 3–16.
8. Budd, G.E., and Jensen, S. (2000). A critical reappraisal of the fossil record of the bilaterian phyla. *Biol. Rev. Camb. Philos. Soc.* 75, 253–295.
9. Fitch, D.H., and Sudhaus, W. (2002). One small step for worms, one giant leap for “Bauplan”? *Evol. Dev.* 4, 243–246.
10. Hall, B. (1999). *Evolutionary Developmental Biology*, second edition (Dordrecht: Kluwer Academic Publishers).
11. Valentine, J., and Hamilton, H. (1998). Body plans, phyla and arthropods. In *Arthropod Relationships*, R. Fortey, and R. Thomas, eds. (Dordrecht: Springer Science+Business Media), pp. 1–9.
12. Wray, G. (2010). Embryos and evolution: 150 years of reciprocal illumination. In *Evolution Since Darwin*, M. Bell, D. Futuyama, W. Eans, and J. Leventon, eds. (Sunderland, MA: Sinauer Associates), pp. 215–240.
13. Kalinka, A.T., Varga, K.M., Gerrard, D.T., Preibisch, S., Corcoran, D.L., Jarrells, J., Ohler, U., Bergman, C.M., and Tomancak, P. (2010). Gene expression divergence recapitulates the developmental hourglass model. *Nature* 468, 811–814.
14. Domazet-Lošo, T., and Tautz, D. (2010). A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* 468, 815–818.
15. Felsenstein, J. (1985). Phylogenies and the Comparative Method. *Am. Nat.* 125, 1–15.
16. Garamszegi, L., ed. (2014). *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology* (Springer).
17. Domazet-Lošo, T., Brajkovic, J., and Tautz, D. (2007). A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* 23, 533–539.
18. Dunn, C.W., Luo, X., and Wu, Z. (2013). Phylogenetic analysis of gene expression. *Integr. Comp. Biol.* 53, 847–856.